

# Proceedings of the Digital Humanities Congress 2018

Lana Pitcher and Michael Pidd

## **Linguistic and Cultural Hegemony in the Digital Humanities**

by Simon Mahony and Jin Gao

### **Citation**

Mahony, Simon, and Jin Gao. 'Linguistic and Cultural Hegemony in the Digital Humanities'. In: Lana Pitcher and Michael Pidd. *Proceedings of the Digital Humanities Congress 2018*. Studies in the Digital Humanities. Sheffield: The Digital Humanities Institute, 2019. Available online at: <<https://www.dhi.ac.uk/openbook/chapter/dhc2018-mahony>>

### **Abstract**

This article aims to form part of wider discussions associated with the development of the digital humanities in the Western sphere and the perceived dominance of the English language and preponderance of text-based scholarship. These discussions intend to raise questions around the lack of engagement that our field has with some geographical regions. As a case study, this article considers mainland China and asks how we might make positive steps to enable closer engagement in order to create a more global digital humanities to the benefit of us all. This study is informed by the first author's networking with Chinese digital humanities research groups and centres, as well as conference attendance in the PRC, and the experiences of the second author as a Chinese graduate student and researcher in the UK.

**1. Introduction** 11 The text of this article was written up following the presentation of the same name given at the [Digital Humanities Congress](#) at Sheffield on 7th September 2018. All the figures, however, were originally presented at the [3rd Peking University Digital Humanities Forum](#): "Incubation and Application: How Digital Humanities Projects Cater to Academic Needs" on 14th June 2018.

Consider the experience of attending a digital humanities conference in a country where you do not have fluency in the language and where there is no live translation or text equivalent. Imagine that the text that follows is the opening introduction, followed by slides with no English and presentations delivered with only the occasional recognisable word such as metadata, big-data, linked-data, XML, TEI (which appear to be part of the international language of digital humanities):

非常感谢大家今天能来听我们关于数字人文多样化的演讲。我在这里用中文作为开场以实践我们研究中所倡导的鼓励学者群多样化的提议。虽然大部分学者都听不懂，但是我希望在不久的将来，更多不同背景不同文化不同语言的学者可以在各种数字人文领域中绽放光彩。 22 Chinese text and the translation at the end of this article are both authored by Jin Gao.

This was the experience of the first author last year at the major digital humanities (DH) conference in the Peoples Republic of China (PRC): the Peking University (PKU) Digital Humanities Forum 2018. This was mitigated by the attention of several Chinese academics and researchers, familiar from other events and networking trips, plus two Chinese graduate level students studying in the UK, one at our own institution and department (the UCL Department of Information Studies), another at our near neighbour in the Strand (the Department of Digital Humanities, King's College London). They must be thanked for their attention, and for making it a memorable experience which will ensure future attendance, although perhaps next time armed with an instant smartphone translation app. The lack of translation at this event was in marked contrast with attendance at high profile library events in the PRC such as Shenzhen 2017: *International*

*Conference on Library and Digital Humanities* and the Shanghai Library: 9<sup>th</sup> *Shanghai International Library Forum* (SILF2018), both of which had live spoken translations via wireless headsets; the former with "whispering translators" for simultaneous interpretation of the parallel sessions in the smaller rooms, and the latter with additional bilingual (English and Chinese) live autogenerated subtitle translations on the large screens located high up on either side of the stage. These library events are clearly more highly funded than the DH Forum, despite the latter being held at one of the top ranked Chinese universities. 33 PKU currently (2019) ranked 30 in the [QS World University Rankings](#) and second to Tsinghua in mainland China. These contrasting experiences fuelled interest and research into diversity and multicultural implications of the digital humanities.

This article will give some background to the development of what has become known as the field of digital humanities, in an attempt to: account for the current situation; raise questions along the way about our concepts of inclusion and diversity; make use of a case study focussing on publications in English and Chinese to identify areas of commonality and difference; and suggest some possible ways in which improvements in diversity (if they are indeed needed) could be facilitated.

## 2. The Ever Evolving Field of Digital Humanities

No area of academic endeavour exists in isolation or emerges without building on what went before. Digital humanities did not spring, like Mithras, from a rock, but has arguably developed and grown out of a movement that already existed within scholarship. This was a movement that brought together researchers and practitioners wishing to progress work and develop methodologies at the intersection of the humanities and technology. We can trace the progression through the changing nomenclature: "applied computing in the humanities", to "humanities computing", and thence to "digital humanities". The purpose here is not to record this development, as that has been done eloquently elsewhere. Nevertheless, as the timing is relevant here, the term itself can be said to date from the early 2000s and was consolidated by the publication of the Schreibman, Siemens and Unsworth's edited volume, *A Companion to Digital Humanities* (Wiley, 2004), and at the same time by the name chosen for the newly launched [ADHO](#), *Alliance of Digital Humanities Organisations*. What became apparent was that what had previously been known as 'humanities computing' had a clear and more complete focus, such as on informatics, building on the traditions of textual and language-based scholarship; 'digital humanities' encompassed a much broader vision for the inclusion of all digital scholarship within the humanities. This wider vision and its possibilities are indeed acknowledged by the authors of the *Companion to Digital Humanities*:

'[...] there are central concerns among digital humanists which cross disciplinary boundaries. This is nowhere more evident than in the representation of knowledge-bearing artifacts. The process of such representation – especially so when done with the attention to detail and the consistency demanded by the computing environment – requires humanists to make explicit what they know about their material [...]. Ultimately, in computer-assisted analysis of large amounts of material that has been encoded and processed according to a rigorous, well thought-out system of knowledge representation, one is connections, and absences that a human being, unaided by the computer, would not be likely to find.' (Schreibman et al, 2004, p. xxvi)

The introductory chapter in Nyhan and Flinn (2016), *Computation and the Humanities: Towards an Oral History of Digital Humanities*, outlines a comprehensive summary of the growth of digital humanities. They acknowledge the difficulties in defining this field, but begin by stating that in their view:

'[...] it takes place at the intersection of computing and cultural heritage. It aims to transform how the artefacts (such as manuscripts) and the phenomena (such as attitudes) that the Humanities study can be encountered, transmitted, questioned, interpreted, problematized and imagined. In doing so it tends to differentiate itself from now routine uses of computing in research and teaching, for example, email and word processing.' (Nyhan and Flinn, 2016, p.1)

The fundamental point here is that, within digital humanities, the computer is not used solely as a tool to aid the scholar (that would be applied computing) but becomes part of the research process itself. It is argued here that the defining element of the digital humanities is that it is a field in which technology and humanistic study come together to mutually benefit from each other with projects that are of intellectual and

research interest to both parties. Neither discipline nor practitioner is the servant of the other, but rather both share in advancing their own specific research agendas so that this partnership facilitates scholarship that would not otherwise be possible. The use of the computer in humanities research, as Busa rightly said way back in 1976:

‘is not aimed towards less human effort, or for doing things faster and with less labour, but for more human work, more mental effort; we must strive to know, more systematically, deeper, and better, what is in our mouth at every moment, the mysterious world of our words.’ (Busa, 1976, p.3)

There is much debate on the competing definitions of the field of DH, and indeed whether or not it constitutes a discipline in its own right; in the future, when everything becomes "digital", will it simply be "humanities", or would that signal the demise of traditional scholarly methods such as close reading and textual criticism? Again, this is a debate that this article chooses not to engage with except insofar as it relates to inclusion rather than exclusion.

Putting constraints on the field, such as requiring practitioners to code and make things (Ramsay, 2011), is at the same time excluding those who do not. Alan Liu (2011) and Geoffrey Rockwell (2011) point to a perceived lack of theory as an instrument of definition. Ironically, the editors of a volume titled *Defining Digital Humanities* (Terras, Nyhan and Vanhoutte, 2013) similarly point to the wide ranging debate, but self-consciously, and in conflict with its title "do not try to define digital humanities [themselves]" but rather "highlight the range of discussions that attempt to scope out the limits and purview of the discipline" (Terras et al, 2013, p.7). It seems reasonable then that definitions, or lack of them, are dependent on areas of practice; if you consider yourself to be a digital humanist then, *de facto*, what you do is assumed (by you at the very least) to be digital humanities. In this way, membership becomes self-referential and indicative of a self-identifying community rather than one that demands any specific practice or skill set (Mahony, 2017). This article seeks to avoid adding to this by taking a pragmatic position: if you say what something is (in this case, a field or discipline of practice) you are also saying what it is not (Mahony, 2018), hence putting up barriers of exclusivity rather than being the inclusive "big tent" that has been suggested and advocated. 44 "Big Tent Digital Humanities", the strapline for the DH211 conference at Stanford University. Nevertheless, in reality, how inclusive are the digital humanities, and to what extent do we represent a community? These are some of the questions to be explored here and the motivation for our presentation at DHC2018.

### 3. Building on What Went Before

The field of digital humanities has grown out of antecedents strongly rooted in linguistic and textual scholarship. The field itself generally looks back to Roberto Busa and his collaborations with IBM to create an *index variorum* of the combined works of Thomas Aquinas, a corpus of medieval Latin texts (Busa, 1980). There are alternative foundational narratives (see Rockwell, 2007; Nyhan and Flinn, 2016); nevertheless, early projects involving the application of computational methods to humanities material came for the most part out of the disciplines of classics and medieval studies (Brunner, 1993; Bodard and Mahony, 2008; Mahony, 2018). Classicists and medievalists were very much at the forefront of humanities scholarship when it came to the use of computational techniques to advance their data-intensive research projects. Since the corpora of ancient sources were often limited, they would have been more manageable, and the scholars themselves were generally trained in the study and interrogation of a variety of source materials and hence more accustomed to the need for combining skills through interdisciplinary and collaborative work practices (Mahony and Bodard, 2010). Nevertheless, these were primarily text-based sources derived from original material found inscribed on stone or written on papyrus, parchment or paper, as well as examples of language and text-based scholarship mentioned above. Moving forward, a cursory glance through recent, self-defined DH publications such as *A New Companion to Digital Humanities* (Schreibman, Siemens and Unsworth eds., 2016), *Debates in the Digital Humanities 2016* (Gold and Klein eds., 2016) and *Digital Humanities* (Berry and Fagerjord, 2017) reveals a much wider field of interest and study to include, among many other facets, modelling, infrastructures, crowdsourcing, knowledge representation, and methods, as well as self-reflective criticism.

There is, however, still a substantial focus on text as an area of research within DH. An analysis by Scott Weingart of the keywords selected from the controlled-vocabulary for uploading proposals for the DH2016

conference, held that year in Krakow Poland, showed that text-related proposals did indeed dominate the declared topics (according to the keywords selected). However, historical studies, text-mining, archives and data visualisation were all on the increase with a new category of "Digital Humanities – Diversity" being added that year (Weingart, 2015). Along with the move away from a focus on text and linguistic-based research, there appears to be a growing interest (or maybe concern) regarding aspects of diversity within the digital humanities more widely.

The issue here with text-based scholarship is that it has grown out of the classical and medieval fields, with those early adopters applying computational methods to Ancient Greek and Latin language sources. In doing so it has, to some extent, reinforced a Western perspective, those being the heritage languages of Europe as we look back to Greco-Roman antiquity as the foundations on which our culture is built; this is particularly so in a geographical region where a classical education was seen as the benchmark of a scholar:

‘Any study of European literature and thought [...] needs to begin with Greece and Rome, and the study of the classics helps to unite the modern man not only with the men of the ancient world but with all those who in later centuries learned from them.’ (Clark, 1959, p.177)

Indeed, Digital Classicists and Digital Medievalists have always been at the forefront of digital humanities research (Bodard and Mahony, 2008; Terras, 2010).

In addition to this, DH has developed in an environment that is dominated, to a large extent, by the English language. Even if our spoken words are not in English, the computer systems that we rely on, with their ones and zeros, respond to and are dominated by the American Standard for Information Exchange (the ASCII code); they display browser pages encoded in HTML with their US-English defined element sets; and transfer data marked-up in the ubiquitous XML with its preference for non-accented characters, scripts that travel across the screen from left to right, and the English language-based TEI guidelines. English is very much the language of the Internet and has become the *lingua franca* of the web. Our domain names are administered by the US based Internet Corporation for Assigned Names and Numbers (ICANN) and currently only available in Latin characters; these are now being extended with the New Generic Top-Level Domains ([ICANN New gTLDs](#)) to include non-Latin characters, although only those that are included in the Anglo/US-centric Unicode. The World Wide Web Consortium (W3C), founded by Tim Berners-Lee at MIT, works on guidelines and standards for the ever-developing web, published in [English language webpages](#). The medium in which we work and correspond has a bias towards the English language and hence towards an Anglophone DH, leading to linguistic differences and regional inequalities (Fiormonte, 2012).

There seems to be similar bias in the favoured language of publication. To have your research read and disseminated widely, there is much pressure to publish in English, which leads to a distortion in the publication metrics. This problem is not limited to the digital humanities but to Western scholarship more generally, as many of the highest rated international journals only publish in English (for example, see the current issue of [Nature](#)). The pressure for publication citations, with the impact of research and its assessment, leads to this hegemony of language. To have your published work widely circulated and read so that it will lead to more citations is essential under the current academic model for academic advancement and promotion. Hence, there is the pressure to publish in English, regardless of native language. The same is often true of major international conferences. Within the field of digital humanities we are seeing a realisation of the bias towards the English language, both in publications and conferences:

‘The over-representation of US and UK Humanities titles [as counted in major indices such as Scopus and Web of Science] will always support arguments in favor of using English as the *lingua franca*, and the misrepresentation of knowledge production and geopolitical imbalance will continue to thrive.’ (Fiormonte, 2015)

This bias appears to be self-perpetuating and, if we are to encourage diversity and the building of wider linguistic academic communities, it needs to be broken. One such initiative is the San Francisco Declaration on Research Assessment ([DORA](#)), which seeks to change the way in which scholarly research is evaluated. Rather than assigning value to the publication venue, with those metrics then being used to evaluate staff and research performance, the research output should be evaluated and assessed according to its own merits. The idea is to avoid what currently seems to be a skewed and biased system that significantly disadvantages early-career researchers and academics from less well-funded institutions in favour of more established



academics and those in prestigious and better funded institutions. The current situation has been aggravated to some extent, rather than relieved, by the move towards Open Access publications, as these often involve an APC (article processing charge), sometimes in addition to the journal subscription already paid for by the institution's library, again favouring better funded research. The same is potentially true for [cOALition S](#) with [Plan S](#) "to make full and immediate Open Access a reality". Nevertheless, DORA is a move towards having a wider range of acceptable publication venues, which will extend the field of options. This initiative to move away from relying on bibliometrics and publication venues as an indicator of the quality of scholarly output is also being supported by major research funders in the UK and Europe such as the European Commission, the (UK) Higher Education Funding Council of England, and the Wellcome Trust. This should go some way to encouraging a wider diversity of publishing venues more generally, which in turn should eventually result in more diversity in the languages of publication.

## 4. Increased Diversity?

As mentioned above, for DH2016 conference proposals the new category of "Digital Humanities – Diversity" was added to the choice of keywords to be attached to proposals to indicate their topic area (Weingart, 2015). Language is not the only limiting factor when it comes to diversity; so, too, is geographical location. The importance of regional diversity was raised by Isabell Galina Russell, based at the Institute for Bibliographic Studies at the National Autonomous University of Mexico (UNAM), in her closing keynote for DH2013, hosted at the University of Nebraska-Lincoln, USA. With the title, *Is There Anybody Out There? Building a Global Digital Humanities Community*, she articulated some of the issues of fundamental concern:

‘One of the things that characterizes DH [...] is that the community has worked very hard towards building the DH community. And most of this work has come from enthusiastic and generous scholars who have given much of their time to developing it. [...] This community has traditionally viewed itself [...] as welcoming and open. Collaboration and cooperation are seen as specific traits of DH [...]. It seems to be that openness and a desire to work with others is fundamental to the way we think of ourselves. And yet, over the past few years this community has become aware that this isn't so open, universal as it thought it was.’ (Galina, 2013)

There is clearly a heightening awareness that there is a need for more community building and diversity within our field. Previously alternating between Europe and North America, the ADHO conference was held for the first time outside this circle in Sydney in 2015, with the professed theme of *Global Digital Humanities*, and acknowledging "the field's expansion worldwide across disciplines, cultures and languages". 55 DH2015, Sydney: <http://dh2015.org> Three years later, DH2018 was hosted at UNAM in Mexico City, which was the first time that the ADHO conference had been hosted in Latin America and the Global South. There is some movement, then, which will be expanded on later in this article.

## 5. Digital Humanities Engagement

From a Western and Anglophone perspective, the major and long-established DH associations and portals, such as ADHO and centerNet, are based in the UK and the USA. Moreover, our major DH journals are predominantly English language publications: *Computers and the Humanities*, *Digital Scholarship in the Humanities* (formerly known as *Literary and Linguistic Computing*), and *Digital Humanities Quarterly*. ADHO incorporates European, Canadian, Australasian, Japanese, French, and Southern African member associations, with a recent new addition, the [Taiwanese Association for Digital Humanity](#) (TADH). Another member association of ADHO is [centerNet](#), "an international network of digital humanities centers formed for cooperation and collaborative action to benefit digital humanities and allied fields [...]". centerNet has a listing of all its members together with them located on a [Google map](#) (unfortunately currently not loading correctly). As viewed (and perhaps inaccurately as it needs attention), it shows a clear preponderance of Western European and North American members, with only a very few outliers. Other than markers for Japan, Taiwan and Hong Kong, the map reveals a distinctly empty space for East Asia. This reflects a similar situation represented in the [infographic](#) map generated from data collected in a survey conducted by Melissa Terras in 2012 to quantify the extent of DH activities globally. In the intervening years little seems to have changed, with one member organisation now missing from South Korea and one added at Hong Kong. East Asia, however, is not the only underrepresented geographical area; other countries that are engaging in DH

research and projects, such as India, are similarly not represented in our global networks. See for example: the [Digital Humanities Alliance of India](#) (DHAI) and the [Centre for Internet and Society](#) (India).

Despite this apparent lack of connection with ADHO or centerNet, and the West more generally, there has been substantial DH-type research activity in mainland China since the early 1990s. The application of computational techniques and methodologies to humanities material in mainland China arguably goes back even further, through the 1970s (the Chinese version of the standard for machine reading of library catalogues, MARC – CNMARC) and the 1980s (Chinese Character Codes for Information Exchange – GB2312-80), followed by the 1990s with [Digital Dunhuang](#), and in the 2000s with CBDB ([China Bibliographic Database](#)) and the [Chinese Text Project](#) (CTP), the latter two based and hosted at Harvard. Just as there is no current definitive "history" of the development of DH in the West (although hopefully forthcoming), there is no such systematic record of these activities in mainland China. 66 The closest we have is Hockey, S., (2004) the History of Humanities Computing, in Schreibman et al (eds.) *Companion to Digital Humanities*. Our second [presentation](#) delivered at DHC2018 breaks this down into four historic periods: Beginning (1979-94 – introduction and digitisation), Development (1995-2001 – database and text search), Consolidation (2002-10 – improvement), and Connection (2011-present – community formation). 77 Hopefully the topic of a post-doc research project focusing on the history and development of DH in mainland China. This is the time, then, one of community formation, to establish connections and form closer relationships with China (as some of us are doing) to make a more global DH community.

This is not, however, the full picture, as there has been a digital humanities centre (the first in mainland China) established at the [University of Wuhan](#), School of Information Management, since 2011 (now back on the centerNet list after an absence, but not indicated with a pin on their map). The University of Nanjing has digital humanities research groups in both the School of History and the School of Arts, as well as the newly formed [Research Centre for Digital Humanities Initiatives](#). Renmin University in Beijing also has a digital humanities research group based in the [School of Information Resource Management](#). These Schools at both Wuhan and Renmin are also members of the [iSchools](#) consortium. These DH groups are all situated within academic faculties, whereas significant other groups in the PRC follow the US, rather than the European, model; at [Peking University](#) (PKU) and the [Shanghai Library](#), the digital humanities research groups are based and work within the context of the library, both with an extensive range of research projects. The institutions mentioned here are ones that have DH research groups and where the first author has been invited to visit to speak and present to both staff and students; there are, without doubt, many others still to connect with.

Globally, the institutional context is not consistent, and so the nature of their activities is not the same; there is no single pattern or model, and hence the research group itself is not a reliable vantage point for establishing similarities and differences:

‘Some centers focus explicitly on digital humanities; some engage the humanities but are organized around media studies, or code studies [...]. North American centers tend to arise from the bottom up, European and Asian centers from the top down. North American centers tend to focus exclusively on humanities and, sometimes, the interpretive social sciences. European and Asian centers are more likely to be dispersed through the disciplines, or to be organized as virtual rather than physically located centers’. (Fraistat, 2012, p. 283)

## 6. Case Study Data Collection and Analysis

The case study used here, because of the differences noted above, considers DH in the PRC specifically by looking at the titles of publications in Chinese and comparing them with those in English in order to identify commonalities and areas of divergence in their topics.

During June 2018, data was collected from two source types to capture the titles of DH journal articles published in both English and Chinese, and separated into two corpora for analytic comparison. The first collected the titles of 3,247 English-language articles taken from the major DH journals over the period 1966 to 2017: *Computers and the Humanities* (CHum), *Digital Scholarship in the Humanities* (DSH) formerly *Literary and Linguistic Computing* (LLC), and *Digital Humanities Quarterly* (DHQ). For the Chinese corpus, searches in Google Scholar for journal articles published during the period from 1964 to 2017 using

the keywords, 数字 + 人文 (digital + humanities) and 人文 + 计算 (humanities + computing) resulted in 1,698 hits.

Another option for searching Chinese publication data for comparative analysis would have been to use [CNKI](#) (China National Knowledge Infrastructure), the largest and perhaps most popular Chinese citation index. This has been used to collect source data for previous bibliometric studies which include information architecture (Lv and Ma, 2010), bibliometric analysis (Qiu, Xiang and Xie, 2010), and co-occurrence networks (Liang, Shi, Tse, Liu, Wang, and Cui, 2009). CNKI allows keyword searching, after which filters can be applied to narrow down the results to help locate relevant publications. The same keyword search, 数字 + 人文 (digital + humanities) and 人文 + 计算 (humanities + computing), however, returned significantly fewer articles (319 and 32 respectively) than the Google Scholar search. This was a surprising result that, although validating our use of Google Scholar to collect the Chinese titles, raises questions as to why that should be, particularly in a region where Google is currently blocked.

In terms of the range of years, this study collected data from as broad a range as possible. For the English publications, all the titles were collected back to the first issues of each journal; thus, this dataset covered the complete publication histories of the three journals until December 2017, when this part of the study was completed. Similarly, the time period was not limited for the Chinese publications. The Chinese dataset included all the publications returned by the keyword searches until June 2018 when the second part of the study was concluded in preparation for the PKU Forum.<sup>1</sup> Although there was a significant difference in the number of individual publications that were collected to form both corpora (3,247 English and 1,698 Chinese), the total number of terms extracted from the publication titles were similar (313 English and 275 Chinese – see below). This appears to reflect the linguistic differences between these two languages, and specifically that the titles of the Chinese DH publications made use of more diverse terminologies (see for example Liang et al, 2009).

Following cleaning and normalisation, both corpora were examined for the words in the titles, the frequencies of those words, and the years of their occurrence. One significant finding in the corpus of Chinese titles was related to the changes in frequency by year. Before 2000, there was very little use of the searched for terms (数字 + 人文 and 人文 + 计算) in the titles, increasing thereafter with greater frequency and then followed by a dramatic rise from 2012 onwards (see Figure 1). Note the spike at about 2004 which coincides with the publication of the Schreibman et al (2004) edited volume, which suggests that these terms were increasing in Chinese publications synchronously with the West.

**Figure 1: The dates of publication of Chinese DH articles taken from Google Scholar using keywords: 数字 + 人文 and 人文 + 计算**

Both corpora were subjected to manual text segmentation and an occurrence count, resulting in 313 English and 275 Chinese unique terms identified in the titles. These are translated and presented here in both languages for comparison in Tables 1 and 2 below.

**Table 1: The 15 most frequent occurrences of Chinese and English title terms**

Chinese terms Occurrences		English terms Occurrences	
1	数字/digital	650	digital/数字
2	图书馆/library	479	analysis/分析
3	研究/the study	225	corpus/文集
			77

4	服务/service	142	language/语言	77
5	建设/construction	117	text/文本	77
6	信息/information	106	digital humanities/数字人文	71
7	资源/resources	104	computing/计算	70
8	发展/development	102	study/研究	64
9	计算/computing	83	using/运用	61
10	分析/analysis	82	english/英语	60
11	云计算/cloud computing	70	literary/文学	60
12	技术/technology	70	based/基于	59
13	模式/mode	59	texts/文本	59
14	论/debate	59	history/历史	57
15	问题/problem	58	database/数据库	53

**Table 2: The 16-30 most frequent occurrences of Chinese and English title terms**

Chinese terms	Occurrences	English terms	Occurrences
16 高校/university	56	project/项目	53
17 应用/application	52	humanities/人文	52
18 我国/china	50	data/数据	50
19 人文/humanities	49	review/评论	49



20 环境/surroundings	48	research/研究	47
21 时代/era	43	introduction/介绍	43
22 管理/management	42	TEI	43
23 探讨/discussion	39	information/信息	40
24 出版/publishing	37	literature/文学	38
25 教学/teaching	36	authorship/作者	37
26 思考/thinking	34	SGML	37
27 体系/system	33	studies/学习	36
28 知识/know how	33	editing/编辑	35
29 新/new	32	case/案例	34
30 档案馆/archives	32	edition/版	34

From this comparison a number of things of interest are apparent. In Table 2, although (with one exception) the terms are all different, the level of the frequency of occurrences are very similar; the terms that are similar, humanities and 人文, also occupy approximately the same position and count. Contrast that with Table 1, where unsurprisingly the highest occurring term is digital 数字 in both sets. The number of occurrences, despite both having relatively similar total number of articles (275 and 313), differs greatly, with the highest having almost five times the number (650 to 135) in the Chinese data compared to English. Indeed, all ten of the most frequently occurring terms are significantly higher in the Chinese corpus. The English corpus must, therefore, have a considerably greater long-tail distribution, allowing for much more variety of combinations.

This difference is more noticeable in the following scatter plots, where the most frequent, earliest and latest terms are shown as examples.

## Figure 2: The occurrences of Chinese terms by the average of the years in which they occur

In Figure 2, the most frequent terms are "digital" and "library". The earliest term, "discussion", appeared in around 1997, while the latest term, "digital humanities" (the two characters 数字人文 used as a single term), appeared in around 2015.

## Figure 3: The occurrences of English terms by the average of the years in which they occur

In Figure 3, the most frequent term is also "digital". The earliest term, "concordance", appeared much earlier than the Chinese term in around 1991, while the latest term, "public", appeared in around 2014.

As can be seen above, the scatter plots of the research topics of the two language publications were significantly different. The Chinese DH-related terms appeared more frequently in a shorter range of years (2000 to 2012), while the English terms were distributed more evenly across the whole period of the timeline. Many of the Chinese terms were used very frequently (e.g., digital, library) with large gaps among other terms, while the English terms were used less frequently but were more evenly distributed over the time period and with smaller gaps between the average year of the occurrences. If we also take into consideration the variation in the total number of Chinese and English publications collected (1,698 and 3,247), the differences between the two languages might be even more significant.

To interrogate this data further, co-occurrence analysis was used to calculate the frequency matrix of any two unique words appearing in the same title. The more times two words appear in the same title, the more closely they are considered to be related. The results were visualised using VOSviewer (1.6.7) showing the node size of each term, the relationships between them in a networked analysis with an average by year in a heat map, and with a year scale included. These are shown below: English terms (with Chinese translations) in Figure 4; Chinese terms (with English translations) in Figure 5; side by side for simple comparison in Figure 6.

**Figure 4: The title word co-occurrence network of English DH publications with data from *CHum LLC/DSH* and *DHQ* (1966-2017)**

**Figure 5: The title word co-occurrence network of Chinese DH publications with data from Google Scholar (1964-2017)**

**Figure 6: The English title network versus the Chinese title network**

An analysis of the figures above is limited to the terms used in the titles of journal publications, rather than the abstract or the content, but nevertheless they are indicative of clear trends. From the overall frequency, the English language title terms have a greater long-tail as well as, from the co-occurrence analysis, a longer history of addressing technical and theoretical topics. They are based more frequently on computational linguistic studies, including those of multiple languages, and address more diverse topics. Conversely, the title terms used in the corpus of Chinese language publications indicated a less diverse range of topics. The most frequently occurring title terms and their combinations account for a much greater proportion of the titles. In addition, there is more emphasis on libraries and tools when compared to the English corpus. Appearing more recently and in contrast with the English corpus, the Chinese title terms indicate a greater interest in publications linked to teaching and its development.

Another limiting factor in the analysis of this case study was the source of the Chinese corpus of articles from which the titles were taken. Although a surprising number of Chinese titles were found in Google Scholar, and that in itself was a significant finding, the number of titles was significantly smaller than those in the English corpus. For a more exhaustive and comprehensive analysis in any future study, we would need to revisit the CNKI database when Chinese DH develops further and has more related journal publications, as currently the results were more significant using Google Scholar. Another future amendment might be to analyse some centrality measures on the two networks to see if that matches our findings here or indicates other possible research directions.

## 7. A More Global Digital Humanities?

From the preceding section, looking at the words in the titles, we can identify the dominant topics of publication in both English and Chinese DH (although limited by our sampling), and through the co-occurrence analysis we can compare the topic clusters to identify points of overlap to statistically identify topics of mutual interest. Where, then, can we make these possible points of contact and shared research interests? An obvious place to start is communication. As noted above, there is a clear spike in the frequency of related Chinese titles at around the time of the publication of Schreibman et al's (2004) edited volume, followed by a dramatic rise from 2012 onwards. Perhaps the communication is only one-way, but a more thorough examination of this potential hypothesis is beyond the scope of this article.

What we can include here are current initiatives to encourage the diversity that would lead to a more global DH and use those as a starting point. As an example, the ADHO DH2018 conference hosted by UNAM in Mexico City, and the first in Latin America and the Global South, with Galina as one of the local organisers, took a more multi-lingual and inclusive approach. With the strapline "[PUENTES/BRIDGES](#)", the call for papers webpage was in German, English, Spanish, French, Italian, and Portuguese, accepting proposals in all these, although limiting the official languages of the conference to Spanish and English. There has, after all, been a well-established DH community in Mexico, [Red de Humanidades Digitales](#) (RedHD), since 2011, with their own Spanish language publication and a clear statement of purpose on their network webpage:

‘Our aims are to promote and strengthen work on the humanities and computing, with special emphasis on teaching and research in Latin America. The RedHD supports better communication between digital humanists in the region [...] the promotion of DH projects [...] the recognition of the field [...] regional projects and initiatives on an international level.’ 88 <http://www.humanidadesdigitales.net/acerca-de>

They too are not listed on centerNet but, since this paper was presented and as of January 2019, they are now a constituent organisation and full member of ADHO. 99 <https://adho.org/announcements/2019/adho-welcomes-new-latin-american-organization>

As discussed above, within DH we are now becoming more self-reflective and are, in some areas, questioning our notions of diversity, particularly so as the dominance of any language is a barrier to inclusivity. ADHO has its own a global initiative for diversity: the [GO:DH](#) (Global Outlook:Digital Humanities) Special Interest Group (SIG). Their online statement defines their purpose:

‘[...] to help break down barriers that hinder communication and collaboration among researchers and students of the Digital Arts, Humanities, and Cultural Heritage sectors in high, mid, and low income economies [...]. Its core activities are Discovery, Community-Building, Research, and Advocacy.’ 1010 <http://www.globaloutlookdh.org/>

GO:DH clearly wishes to encourage, and perhaps facilitate, communication and the collaboration that may result, particularly between regions whose economies are of a different level, the haves and the have-nots. In addition, scrolling to the bottom of their pages (and they might have foregrounded these more to make the point more strongly), there are links to pages in English, Chinese (both Simplified and Traditional), French, Italian, Spanish, Arabic, Nepali, and Brazilian Portuguese. At the time of writing, not all of these were working; nevertheless, they are taking a welcomed lead in promoting linguistic diversity. In addition, ADHO has a Multi-lingualism and Multi-culturalism Committee ([MLMC](#)), with a wide range of nationalities represented along with policy and objectives statements to help members "to become more linguistically and culturally inclusive in general terms, and especially in the areas where linguistic and cultural matters play a role". These initiatives show a clear awareness of the need to be more inclusive within DH with regards to diversity of both culture and language.

These initiatives are, indeed, laudable, but perhaps there is more that we should be thinking about when it comes to diversity – perhaps our structures as well. In the West we might put ADHO, their conferences, and our preferred publication models and venues at the centre of DH, but that is looking at things from our perspective and applying our cultural (and academic) biases. Potential colleagues and collaborators in other parts of the globe perhaps see things differently. Think about the traditional school atlas and how that inculcates a particular view of the world; they are different in different global regions for good reasons.

There are alternatives to a Western/Anglophone centric view of DH which places others at the periphery as outliers, and Amy Earhart, herself a member of GO:DH, suggests the following:

‘Instead of insisting that we encapsulate all practices of digital humanities within a big tent or a centralized structure, we should instead view ADHO and its conferences and journals as important, but not central, meeting spaces for digital humanists. Rather than seeing ADHO as the center, we should encourage a global digital humanities that works on the borderlands, with localized expressions of scholarship that reinvigorate through exchange. [...]

[This] is the only way that we might move beyond binaries that are currently in place, whether technologically advanced/primitive, east/west, or low income/high income. Resisting the homogenization of scholarly methods, questions, outcomes, production and ownership is the only way to develop a truly robust global digital humanities.’ (Earhart, 2018, p. 368) 1111 This publication resulted from a conference at Fudan University in 2017, "Cross-cultural, Cross-group and Comparative Modernity Conference", which the first author also attended and where we both took the theme of diversity for our presentations. Both articles are in the same print and online journal, Mahony (2018) and Earhart (2018) but whereas the former is available freely as Open Access, the latter is only available on subscription or as a pre-print in the institutional repository.

## 8. Conclusions

Pulling everything together, there has been substantial Chinese DH-type research ongoing in mainland China, with an established centre at Wuhan since 2011, others currently being established, and active DH research groups in several libraries including Shanghai and PKU. Despite this, there are no apparent connections between them and organisations such as ADHO and centerNet (with the one exception noted above). There are individual connections being established on a personal level and, of course, the Harvard connections with CBDB and the CTP via the Fairbank Centre for Chinese Studies.

The extent of DH projects in the PRC is considerable, although much of it is unknown in the West. [Wuhan](#) and [Nanjing](#) are developing their online DH presence and PKU library are planning their DH research projects website. The DH group at Shanghai Library has a significant amount of content online, including material released freely under Creative Commons license: [Chinese Genealogy Knowledge Service Platform](#); [Chinese Ancient Books Union Catalogue and Evidence-based Platform](#); and a [Knowledge Base](#) hub to pull all their projects together. Figures 7, 8 and 9 are the flyer, plus recto and verso, of the English language version of their research projects brochure produced for the [SILF2018](#) conference.

**Figure 7: The promotional flyer listing Shanghai Library DH projects with URLs and QR codes**

**Figure 8: The recto of the Shanghai Library DH projects English language brochure**

**Figure 9: The verso of the Shanghai Library DH projects English language brochure**

There are clear language obstacles which should not be underestimated, but these appear, on anecdotal evidence only, to be one-sided, as many Chinese academics are able to read and speak English. If so, they are able to access English language publications, which may account for the spike in the Chinese DH-related titles in Figure 1 above, whereas those of us who only read English (and/or other European languages) have very limited access to the content of research publications in Chinese. There are also publication differences relating to venues as well as language and the availability of open access publishing. Funding bodies should be encouraged to consider the benefits of investing in the translation of research output for wider



dissemination, readership, and subsequently the promotion of their own institutions. The same is true of conference attendance and the success of having proposals accepted at events with limited language requirements. Just as at DH2018 at UNAM, we should think about allowing more diversity of languages for conference proposals; consider the difficulties in getting a proposal accepted in a language other than your own. Once there, translation facilities need to be considered as part of the package for conference funding.

As well as language difficulties there are also cultural barriers, including our perhaps too Western-centric view and how that colours our perceptions. To broaden our own perspectives we might shift our thinking away from ideas of a centralised DH structure, as suggested by Earhart, with its "distinct bias towards North American and European notions of culture, value and ownership" (Earhart, 2018, p. 357), and embrace diversity through more regional networks of what she refers to as "localized borderlands". Perhaps, after all, "the region is less important than other forms of constituency as an organising principle" for the digital humanities (O'Donnell et al, 2016, p. 500). The developing relationships between individuals in different geographic regions and their personal networks may well be a better model.

The format of relationships within research groups and centres, regardless of whether they are in the faculty or the library, are all different; there is no one single model. Whatever the arrangements:

‘[...] the mission will be [...] to build greater connectivity and collaboration between and across existing centers, resources, and practitioners [...]. In pursuing that mission, building and creating networks is the most important activity of all.’ (O'Donnell et al, 2016, p. 473)

To a great extent, this ability to make connections, establish relationships, and create networks is dependent on institutional and financial constraints. Without communication there can be no exchange of the ideas that might lead to collaboration, which is the essential constituent in digital humanities research and practice, and the seeding of future projects. This relies heavily on the availability of travel budgets and funds to host visiting scholars and researchers, as well as conference attendance, but also on our willingness to step outside of our familiar comfort zones to reach out to new audiences and to engage beyond our limited echo-chamber. Restricting linguistic and cultural perspectives restricts our field, whereas inclusion benefits us all.

## 9. Coda: translation of the Chinese introduction

Thank you very much for coming to our talk about the diversity of digital humanities.

I am here to use Chinese as an introduction to practice our proposal to encourage the diversity of scholars in digital humanities. Although most scholars do not understand this, we hope that in the near future more scholars from different backgrounds, cultures, and languages will be able to shine in various digital humanities fields.

## 10. References

Berry, D. and Fagerjord, A. (2017) *Digital Humanities: Knowledge and Critique in a Digital Age*, Polity Press.

Bodard, G. & Mahony, S eds. (2008) ""Though much is taken, much abides": Recovering antiquity through innovative digital methodologies", Digital Classicist special issue, *Digital Medievalist* 4.  
<<https://journal.digitalmedievalist.org/articles/10.16995/dm.17>> [Accessed 25/02/2019]

Bodard, G. & Mahony, S. eds. (2010) *Digital Research in the Study of Classical Antiquity*, Ashgate

Brunner, T. F. (1993). Classics and the computer: The history of a relationship. In J. Solomon (eds.), *Accessing antiquity: The computerization of classical studies* (pp. 10–33). Tucson: University of Arizona Press.

Busa, R., (1976) Why can a computer do so little? *ALLC Bulletin* 4.1,1-3.

- Busa, R., (1980), "The Annals of Humanities Computing: The Index Thomisticus." *Computers and the Humanities* 14: 83–90.
- Clarke, M. (1959) *Classical education in Britain 1500-1900* Cambridge University Press.
- Earhart, A. (2018) "Digital Humanities Within a Global Context: Creating Borderlands of Localized Expression", *Fudan Journal of the Humanities and Social Sciences* 11: 357–369. Springer
- Fiormonte, D. (2012). "Towards a Cultural Critique of the Digital Humanities", *Historical Social Research*, Vol. 37 No. 3, 59-76. Reprinted in: *Debates in the Digital Humanities 2016*, eds Matthew Gold and Lauren Klein (eds.) University of Minnesota Press.
- Fiormonte, D. (2015). 'Towards monoculture (digital) humanities?' *Infolet: Cultura e critica dei media digitali* :7. <<https://infolet.it/2015/07/12/monocultural-humanities>> [Accessed 25/02/2019]
- Fraistat, N. (2012), "The function of digital humanities centers at the present time", in *Debates in the Digital Humanities*, ed. Gold, M., University of Minnesota Press.
- Galina Russell, I., (2013) "Is There Anybody Out There? Building a global Digital Humanities community", *Humanidades Digitales* <<http://humanidadesdigitales.net/blog/2013/07/19/is-there-anybody-out-there-building-a-global-digital-humanities-community>> [Accessed 25/02/2019]
- Gold, M. and Klein, L. eds. (2016): *Debates in the Digital Humanities 2016*, University of Minnesota Press.
- Liang, W., Shi, Y., Tse, C. K., Liu, J., Wang, Y., & Cui, X. (2009). "Comparison of co-occurrence networks of the Chinese and English languages". *Physica A: Statistical Mechanics and Its Applications*, 388(23), 4901–4909. <<https://doi.org/10.1016/j.physa.2009.07.047>> [Accessed 25/02/2019]
- Liu, A., (2011) "Where is the Cultural Criticism in the Digital Humanities", in Gold, M. (eds.) *Debates in the Digital Humanities*, University of Minnesota Press: 490-509
- Lv, H., & Ma, H. (2010). "Bibliometric Statistical Analysis and Evaluation on Information Architecture Research in China in the Past 8 Years". *Information Science*, 2010(10). Retrieved from <[http://en.cnki.com.cn/Article\\_en/CJFDTotal-QBKX201010021.htm](http://en.cnki.com.cn/Article_en/CJFDTotal-QBKX201010021.htm)> [Accessed 25/02/2019]
- Mahony, S. and Bodard, G. (2010) "Introduction", in Bodard & Mahony (eds.) *Digital Research in the Study of Classical Antiquity*, Ashgate
- Mahony, S. (2017). "The Digital Classicist: building a Digital Humanities Community", *Digital Humanities Quarterly* 11:3. <<http://www.digitalhumanities.org/dhq/vol/11/3/000335/000335.html>> [Accessed 25/02/2019]
- Mahony, S. (2018) "Cultural Diversity and the Digital Humanities", *Fudan Journal of the Humanities and Social Sciences* 11: 371-388. Springer. <<https://link.springer.com/article/10.1007/s40647-018-0216-0>> [Accessed 25/02/2019]
- Nyhan, J. and Flinn, A. (2016), *Computation and the Humanities: Towards an Oral History of Digital Humanities*, Springer.
- Qiu, D., Xiang, P., & Xie, G. (2010). "Bibliometric analysis on published papers authored by staffs in Sichuan Academy of Agricultural Sciences based on CNKI literature database". *Information Science*, 2010(8). Retrieved from <[http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-LYTS201008029.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-LYTS201008029.htm)> [Accessed 25/02/2019]
- Ramsay (2011) *Who's In and Who's Out*, Lenz.uni.edu.
- Rockwell, G. (2007) "An alternate beginning to humanities computing?" *Theoretica.ca*.
- Rockwell, G. (2011) "Inclusion in the Digital Humanities", *philosophi.ca*. <<https://philosophi.ca/pmwiki.php/Main/InclusionInTheDigitalHumanities>> [Accessed 25/02/2019]

Schreibman, S., Siemens, R., and Unsworth, J. (eds.) (2004) *A Companion to Digital Humanities*, Blackwell.

Schreibman, S., Siemens, R., and Unsworth, J. (eds.) (2016) *A New Companion to Digital Humanities*, Wiley-Blackwell.

Terras, M. (2010) "The Digital Classicist: Disciplinary Focus an Interdisciplinary Vision", in Bodard & Mahony (eds) *Digital Research in the Study of Classical Antiquity*, Ashgate

Weingart, S. (2015). "Submissions to DH2016 (pt. 1)", *The scottbot irregular*.  
<<http://www.scottbot.net/HIAL/index.html@p=41533.html>> [Accessed 25/02/2019]

,

## Endnotes

1 The text of this article was written up following the presentation of the same name given at the [Digital Humanities Congress](#) at Sheffield on 7th September 2018. All the figures, however, were originally presented at the [3rd Peking University Digital Humanities Forum](#): "Incubation and Application: How Digital Humanities Projects Cater to Academic Needs" on 14th June 2018.

2 Chinese text and the translation at the end of this article are both authored by Jin Gao.

3 PKU currently (2019) ranked 30 in the [QS World University Rankings](#) and second to Tsinghua in mainland China.

4 "Big Tent Digital Humanities", the strapline for the DH211 conference at Stanford University.

5 DH2015, Sydney: <http://dh2015.org>

6 The closest we have is Hockey, S., (2004) the History of Humanities Computing, in Schreibman et al (eds.) *Companion to Digital Humanities*.

7 Hopefully the topic of a post-doc research project focusing on the history and development of DH in mainland China.

8 <http://www.humanidadesdigitales.net/acerca-de>

9 <https://adho.org/announcements/2019/adho-welcomes-new-latin-american-organization>

10 <http://www.globaloutlookdh.org/>

11 This publication resulted from a conference at Fudan University in 2017, "Cross-cultural, Cross-group and Comparative Modernity Conference", which the first author also attended and where we both took the theme of diversity for our presentations. Both articles are in the same print and online journal, Mahony (2018) and Earhart (2018) but whereas the former is available freely as Open Access, the latter is only available on subscription or as a pre-print in the institutional repository.